

A Consistency Test for Phylogenies Based on Contemporaneous Species

EDWARD O. WILSON

It is commonly stated that phylogenies deduced from data about contemporaneous species cannot be "proved" because, obviously, evolution is a past event recoverable only from fossils. This is true to the extent that no proof exists which has the decisiveness of a witnessed event or the consistency of a physical measurement. Yet scientific proof is rarely direct and is always relative in degree. Evolutionary hypotheses might never be definitive by the standards of experimental biology, but they are valid if they are both falsifiable and heuristic. That is, to be valid they should make concrete predictions that are capable of being negated if the hypothesis is false; and they should point the way to deeper, more meaningful investigations if they are momentarily upheld. Phylogenetic taxonomy has been open to criticism not so much for indirection as for its lack of techniques of formal analysis that render its hypotheses falsifiable and heuristic.

One such procedure that might be employed involves the "weighting" of characters with reference to their phylogenetic significance. Taxonomists intuitively select character states which they postulate to define monophyletic sets of species. The ideal character contains some state that both uniquely defines a set of species and has not been reversed in evolution, so that all existing species which possess this state can be said to have descended from one species in the past that evolved the state. For every such character state that can be identified, a branch in the phylogenetic tree can be added. This extreme form of phylogenetic hypothesis, then, is initiated as a hypothesis about unique, unreversed characters. The formulation perhaps cannot be decisively

proven on the basis of contemporaneous species. But can it be disproven? And is anything of biological significance to be gained by the procedure? The following test hopefully gives an affirmative answer to both questions. It is not original in the sense that it offers something very new to taxonomic thinking. Instead, its purpose is to express one common intuitive taxonomic procedure in a new, more rigorous form.

Definitions. Consider a series of m unique, unreversed character states $a_1, a_2, a_3, \dots, a_m$ each representing a different character, as yellow dewlap and flattened tail can be said to be states of two separate characters (dewlap color and tail shape) in lizard species. These particular states are interpreted to have appeared during a speciation episode that has resulted in a monophyletic taxon of n contemporaneous species. They now exist in any combination in various of the n species. At one extreme, they may be totally lacking in a given species; at the other extreme, all m character states may occur together in a given species. By *unique* is meant that a given character state a_i appeared in the past only once and in one species. It now exists in one or more descendant species. By *unreversed* is meant that the state has never been lost, i.e., has never reverted to a prior state, in any of the species giving rise to the contemporaneous taxon. The character state itself might have arisen *de novo* as a new structure, it might have appeared as a new state in a series of discrete character states, or it might be arbitrarily recognized as some point and beyond in a continuous morphocline. The taxon therefore is to be treated as a sample space whose points are contemporaneous species; and the character states are events

that can occur on the points in the sample space. It is desirable that the character states in the hypothesis be chosen initially for considerations not having to do with the way they jointly define sets of species in the taxon. The properties that can be expected to induce the choice include uniqueness with reference to other taxa, structural complexity, and absence of other states that are clearly annectant or derivative and degenerate in nature.

The hypothesis. The m character states are unique and unreversed.

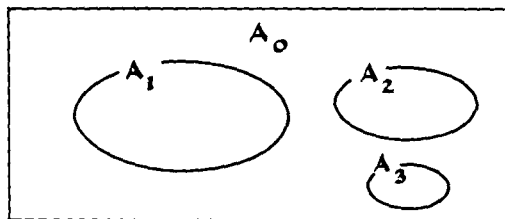
*Testing the hypothesis.*¹ Let us label the states such that the possession of a_1 completely defines a set of contemporaneous species A_1 , a_2 defines a smaller or equal set A_2 , a_3 defines a still smaller or equal set A_3 , and so on. Three possible alternative outcomes can now be simply stated (Fig. 1).

I. If the sets of species defined by the character states are non-overlapping, i.e., $A_1 \cap A_2 \cap A_3 \cap \dots \cap A_m = 0$, the hypothesis cannot be tested.

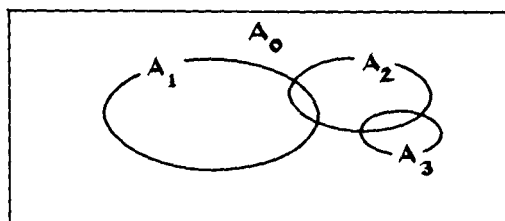
II. If the sets are overlapping but do not enclose each other (form a series of proper subsets), in the order $a_1, a_2, a_3, \dots, a_m$, the hypothesis is rejected.

III. If the sets are overlapping and A_2 is wholly enclosed in (is a proper subset of) A_1 (and A_3 is enclosed in A_2 , and so on) the arrangement is consistent with the hypothesis but does not definitely prove it. If Situation III holds, the following phylogenetic hypothesis is also consistent: $A_0 \cap \bar{A}_1, A_1 \cap \bar{A}_2, \dots, A_{m-1} \cap \bar{A}_m, A_m$ are the contemporaneous branches of a phylogenetic tree of the kind illustrated in Figure 2.

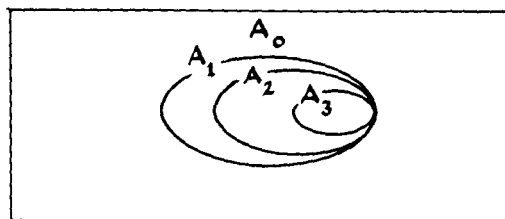
Examples of reasonably long sequences of character states that pass the consistency test are probably familiar to most taxonomists. Two such sequences from the ants



I. Test not applicable



II. Test failed, hypothesis rejected



III. Test passed, hypothesis not rejected

FIG. 1. The consistency test. The rectangle encloses the taxon under consideration. It must be reasonably discrete in many characters from all other taxa. Each ellipse encloses a set of species A_i characterized by a character state a_i hypothesized to be unique and unreversed. Only one state per character is considered. A_0 is the set of species not bearing any character state hypothesized to be unique and unreversed.

¹ The following conventions from set theory are used: A_1 symbolizes the set of all species that possess character state a_1 . \bar{A}_1 symbolizes the set of all species that do not possess character state a_1 . $A_1 \cap A_2$ indicates those species that are in both A_1 and A_2 , i.e., possess both character states a_1 and a_2 . $A_1 \supset A_2$ means that A_2 is contained wholly within A_1 , i.e., all species that possess a_2 also possess a_1 .

(family Formicidae) are given in Table 1.

Suppose that the m character states are mutually consistent, as in Situation III. Although it is not possible from this fact alone to prove the phylogenetic hypothesis of Figure 2, we might still be able to narrow

the permissible alternative explanations somewhat. First, consider the model which is the opposite of the one under consideration, namely that the m states have appeared and disappeared in a random manner with reference to each other during the evolution of the taxon. This hypothesis can be tested in the following manner. Imagine the circumstance, among all possible circumstances, in which there would exist the highest probability of the nested pattern arising by chance combinations alone. This is the simple case illustrated in Figure 3. There are $m + 1$ species evolving separately during the time that the characters are fixed (at random with reference to each other) to produce the nested pattern. Given that at the time m of the species acquired a_1 , $m - 1$ acquired a_2 , $m - 2$ acquired a_3 , *et seq.*, with a single species acquiring a_m , the probability that the resulting sets A_i could be nested by chance alone is

$$P(A_1 \supset A_2 \supset \dots \supset A_m) = \frac{[m!2!][(m-1)!3!]\dots[3!(m-1)!][2!m!]}{[(m+1)!]^{m-1}}.$$

All other situations in which the character states are fixed independently to give a nest of sets are equally or less probable. In other words, the equation above, based on the random model illustrated in Figure 3, gives an upper limit for the probability that the m character states were evolved randomly with reference to each other. Applying it for various values of m we find that $P(A_1 \supset A_2 \supset \dots \supset A_m)$ is 6/125 for four character states, 1/225 for five character states, 16/84,035 for six character states, and 9/153,664 for seven character states. In order for this explicit formulation to be valid it is necessary that the character states be chosen initially without reference to the kind of classification they would engender in the taxon. In practice such selection would come about in the first study of a group of species, before the distribution of various character states with reference to each other are considered.

In sum, if four or more character states hypothesized to be unique and unreversed

Sets of Contemporaneous Species

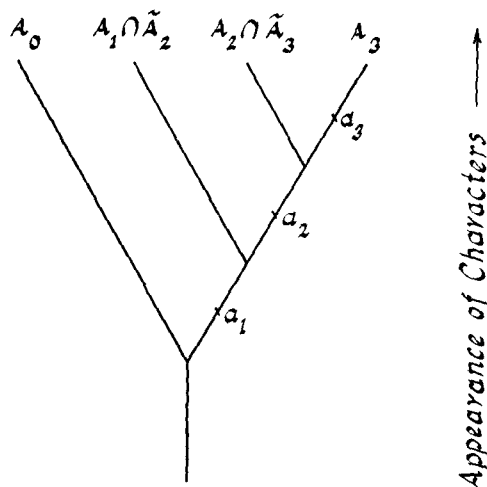


FIG. 2. The phylogenetic hypothesis (cladogram) that is permissible if the character states a_1 , a_2 , and a_3 pass the consistency test. Each character state represents a *different* character. A_0 , $A_1 \cap A_2$, $A_2 \cap A_3$, and A_3 represent sets of contemporaneous species. The nodes labeled a_1 , a_2 , and a_3 mark the appearance of these character states in time. The ends of the branches are arbitrarily arranged along equal intervals because the consistency test by itself gives no information about over-all similarity of the sets of species.

then pass the consistency test, we are reasonably justified in considering them correlated in some historical manner, regardless of the pathways of speciation taken by the taxon in the past. With much more confidence, this rule can be based on five or more characters.

Suppose the consistency test is thus passed with reasonable confidence. There are five alternative ways in which the character states could be correlated:

1) The m character states were fixed at random. Later, there was differential survival among the species according to their respective combinations of the m states, resulting in the modern consistent pattern. Unless we also postulate genetic drift, this explanation subsumes that the ways that the m states are combined are at one time selectively neutral and later selectively significant. The probability certainly exists

TABLE 1—SEQUENCES OF INTERCONSISTENT CHARACTER STATES IN THE FORMICIDAE.

	Character State	Group Defined within the Taxon
Series No. 1 (Taxon = Aculeate Hymenoptera)	Metapleural gland	Family Formicidae
	Pulvinate poison gland	Subfamily Formicinae
	Sepalous proventriculus	"Section Euformicinae"
	Dense, appressed pilosity in discrete soldier caste	Subgenus <i>Machaeromyrma</i> of <i>Cataglyphis</i>
Series No. 2 (Taxon = Genus <i>Lasius</i>)	"Niger-type" male mandible	<i>Lasius</i> exclusive of Subgenus <i>Chthonolasius</i>
	Metapleural guard hairs lost in female castes	Subgenus <i>Dendrolasius</i>
	β -form queen	<i>L. teranishii</i> and <i>L. spathepus</i>
	Appendages covered with long, silvery pilosity	<i>L. spathepus</i>

but seems intuitively relatively small. Or,

2) A superordinate character state, e.g., a_1 with reference to a_2 , always or with very high frequency appears soon after the subordinate appears; but it also originates in a certain fraction of the species without the subordinate character state. This possibility seems even more remote than (1). Or,

3) A subordinate state, e.g., a_2 with reference to a_1 , occurs only after the superordinate state is present. But it can still be nonunique and reversible within the species bearing the superordinate state. For example, a_2 could still appear and disappear many times over in species bearing a_1 . This is perhaps more likely than (1) and (2); however, if the subordinate character states really could appear and disappear in multiple fashion within the set of species bearing superordinate character states, it would be necessary for each of the subordinate states to have changed in concert to preserve the consistency observed in the contemporaneous taxon. Or,

4) The character states could first have appeared together and then been lost in concert to produce the precise pattern. Or,

5) The states are unique and unreversed within the taxon. This seems the most likely hypothesis. It is certainly the simplest.

Heuristic value. Consistent phylogenetic schemes, even when based entirely on con-

temporaneous species, are useful for two reasons: they serve to confirm the identity of the most unusual and stable character states, and they make exact predictions about state combinations in the species yet to be discovered. While remaining explicitly vulnerable, they are a valuable scientific procedure, comprising that part of taxonomic research which has the greatest general interest. This positive aspect of phylogenetic analysis holds whether or not enough characters pass the consistency test to allow the random hypothesis to be rejected. It also holds whether or not the phylogeny deduced is correct in detail and regardless of its effect on formal classification.

An example illustrating the heuristic value of cladistic analysis can be taken from my recent revision of the ant genus *Aenictus* (Wilson, 1964). Two character states, the "Typhlatta spots" of the head and presence of teeth on the anterior margin of the clypeus, were among those initially guessed to be unique and unreversed, but they proved not to be interconsistent. In particular, *Aenictus currax* and *A. huonicus* possessed Typhlatta spots but appeared to lack clypeal teeth. Since these two species are very similar in all characters studied, they were inferred to be closely related. Also, since they are both endemics of New Guinea,

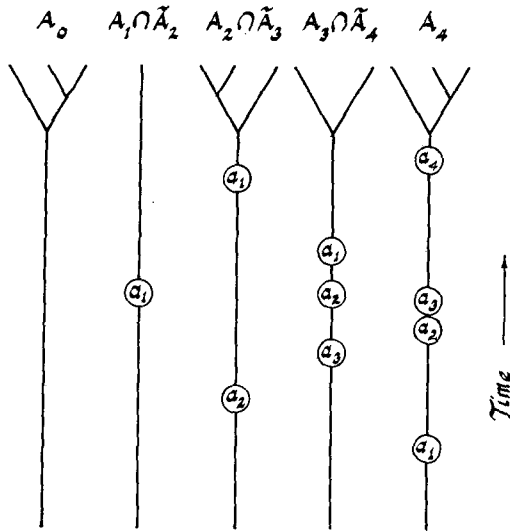


FIG. 3. The situation in which there would be the highest probability of m character states (in this case $m = 4$) passing the consistency test while evolving in a random manner. There are $m + 1$ species during the appearance of the states. The placement of the numbers indicates the appearance of the states in time. They are scattered arbitrarily in this diagram to suggest the condition of randomness. All other situations would give equal or higher probabilities. After the states appear the $m + 1$ species may or may not speciate further to produce the contemporaneous taxon, as exemplified by the irregular branching near the ends of the phyletic lines.

which lies on the periphery of the range of the species group to which they belong, it was guessed that any deviant character states shown by them would be more likely to be derived than original. This second deduction was based on a rule shown by Indo-Australian ant species generally. A closer, second examination of the *Aenictus* species resulted in support for the hypothesis: workers of *A. currax* were found to have hidden, rudimentary teeth. As a consequence, it was concluded that clypeal teeth have been lost secondarily in *A. huonicus*. It is my impression that similar logical sequences are often, even routinely followed in taxonomic revisions. Taxonomists seldom spell their procedures out, however, as I have done in the *Aenictus* revision.

Relation to classification. Formal classifications need not be isomorphic with phylogenetic schemes that are simply cladistic in nature. The sets of species $A_i \cap \bar{A}_{i+1}$ defined by characters that continue to pass the consistency test may or may not be recognized as taxa. It is conceivable, for example, that a species in A_2 differs from one in $A_1 \cap \bar{A}_2$ only by the character state a_2 but is different from other species in A_2 by many other characters. In this case it would be valid taxonomic procedure either to lump $A_1 \cap \bar{A}_2$ and A_2 or to combine the one species from A_2 with $A_1 \cap \bar{A}_2$; and it would be dubious procedure to split $A_1 \cap \bar{A}_2$ and A_2 as taxa. This conclusion has been reached by members of both the phylogenetic (Simpson, 1961) and numerical schools (Sokal and Sneath, 1963).

Even so, the present study together with independent and parallel attempts to formalize cladistic analysis (e.g., the articles by Sokal and Camin and by Throckmorton in this issue of *Systematic Zoology*) indicate that we can hope to distinguish with confidence between "constant characters" and "fickle characters." In the aggregate, constant characters reflect phylogenies more accurately than fickle characters, and it would appear that insofar as we wish to transmit evolutionary information in our classifications constant characters should be given greater weight. Such classifications may not be as stable and reproducible as those based on the averaged similarity of unweighted characters, but they have more biological interest. Taxonomy should be more than the blind clustering of taxa according to over-all similarity, as suggested by the "numerical taxonomists." In spite of the attractive simplicity of the latter technique and its undoubted usefulness in special cases, it seems to be of dubious value as a broad philosophy of classification. The main objection is that numerical taxonomy has up to the present offered little hope of yielding new biological information, precisely because it has not been constructed with reference to any real biological questions. Put in another way, taxonomy is a

language that can be designed according to any one of many sets of rules. The rules selected should be of maximum heuristic value; beyond that, it is only necessary that they be stated very plainly. While not intending to disparage multivariate statistics or the considerable technical achievements of the numerical taxonomists, I would regard a taxonomy based automatically and *a priori* on unweighted characters as a desirable measure only in cases where phylogenetic hypotheses cannot in any way be tested. Even at its best this procedure should never be accepted as a doctrine. In fact, it seems more likely than ever before that taxonomists will eventually develop standard methods for the combination of phenetic measures and cladistic inferences into truly phylogenetic classifications. To do so would be one of the great achievements of modern evolutionary biology.

Summary

Cladograms of contemporaneous species are most rigorously constructed from a hypothesis which postulates unique, unreversed character states (Fig. 2). Many such phylogenetic schemes, if false, can be quickly discarded by a simple consistency test illustrated in Figure 1. If a set of four or more character states $a_1, a_2, a_3, \dots, a_m$ found in m different characters in a taxon are selected without reference to the grouping of species within the taxon and then are found to characterize successively smaller sets of species in such a way as to pass the consistency test, it is reasonable to conclude that the character states evolved in the taxon in a non-random manner with respect to each other. The relation of this inference to phylogeny and the heuristic value of phylogenies based solely on contemporaneous species are discussed.

Acknowledgments

I am very grateful to Eli Minkoff and Angelo Serra for critical readings of the manuscript. Several other persons, including A. F. Bartholomay, W. H. Bossert, W. L. Brown, H. E. Evans, R. Inger, E. MacLeod, E. Mayr, G. G. Simpson, R. R. Sokal, and

R. W. Taylor, have discussed various aspects of the problem and provided help and encouragement. The consistency test was developed in conjunction with a recent systematic study (Wilson, 1964) of the Indo-Australian doryline ants supported by a grant from the National Science Foundation.

REFERENCES

- SIMPSON, G. G. 1961. Principles of animal taxonomy. Columbia University Press, New York, 247 p.
 SOKAL, R. R., and P. H. A. SNEATH. 1963. Principles of numerical taxonomy. W. H. Freeman, San Francisco, 359 p.
 WILSON, E. O. 1964. The true army ants of the Indo-Australian area (Hymenoptera: Formicidae: Dorylinae). Pacific Insects 6:427-483.

Appendix

The following is a proof of the proposition that the model in which $m \geq 3$ character states are fixed in $m + 1$ species gives the highest probability, among all possible models, that the m character states could have evolved at random with respect to each other and still have been fixed interconsistently. Let the array of numbers of species in each group A_i vary and any given array be labeled with a number j ; in the extreme case $j = 1$, there exists the extreme model just cited in which the number of species in A_0 is $m + 1$, the number in A_1 is m , the number in A_2 is $m - 1$, and so on. In short, the array $j = 1$ contains the smallest number of species possible. In a second array $j = 2$, A_0 might contain $m + 2$ species, A_1 m species, A_2 $m - 1$ species, and so on. The probability that a given array occurred as the character states were fixed is p_j and $\sum_j p_j = 1$.

The probability that in an array j the m character states would be fixed in a given pattern with respect to one another can be designated q_{jk} . In particular let us label as $k = \alpha$ the condition in which the character states turn out to be interconsistent. What is desired is the maximum value of $\sum_j p_j q_{j\alpha}$ for all possible arrays j . Now it is intuitively apparent and has been borne out by inspection of many cases (but not

formally proved for all possible cases) that where $m \geq 3$ the maximum value of $q_{j\alpha}$ is obtained when the array j contains the smallest number of elements, i.e., in the case $j=1$. The maximum value for $q_{j\alpha}$ is $q_{1\alpha}$, which is a constant when m is chosen. Consider any array j that occurred in evolution. Then given some value of m ,

$\sum_j p_j q_{j\alpha} \leq \sum_j p_j q_{1\alpha}$ for all j and, since $\sum_j p_j q_{1\alpha} = q_{1\alpha} \sum_j p_j$ for the special "alpha case" and $\sum_j p_j = 1$, it follows that $\sum_j p_j q_{j\alpha} \leq q_{1\alpha}$.

EDWARD O. WILSON is in the Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138.